

Statistics for Economists II

Week 3: Point Estimation — Bias, MSE & Efficiency

Güneş A. Aşık

Department of Economics
Middle East Technical University

Spring 2025–2026

Roadmap for Today

- 1 Estimator, Estimate, and Estimation
- 2 Unbiasedness
- 3 The Dartboard Analogy: Bias and Variance Visualized
- 4 Efficiency
- 5 Mean Square Error: The Bias–Variance Tradeoff
- 6 Consistency
- 7 Choosing the Best Estimator: Uniform Distribution Example
- 8 Practice Problems
- 9 Key Takeaways

The Big Question of Statistics

We observe a **sample** $X_1, X_2, \dots, X_n$ from a population with unknown parameter θ .

Goal: Use the sample to “guess” the value of θ as accurately as possible.

This chapter answers: *What makes a good guess?*

Examples

- What is the **average wage** in Turkey?
- What is the **true unemployment rate**?

We never observe the entire population—we **estimate** from samples.

Three Distinct Concepts

Estimator, Estimate, and Estimation

Estimator: A **rule** (function of the sample) used to guess θ .

It is a *random variable*: $\hat{\theta} = T(X_1, \dots, X_n)$

Estimate: The **numerical value** obtained when the rule is applied to a specific sample.

It is a *number*: $\hat{\theta} = T(x_1, \dots, x_n) = 4.73$

Estimation: The entire **process** of choosing and applying an estimator.

Think of it this way:

- **Estimator** = the *recipe* (e.g., “take the sample mean”)
- **Estimate** = the *dish* from one batch of ingredients
- **Estimation** = the act of *cooking*

Since it is a random variable, it has a distribution, mean, variance—we can judge how “good” it is.

Example: Estimating Average Monthly Wage

Suppose monthly wages in a city follow some distribution with unknown mean μ and variance σ^2 .

We draw a random sample of $n = 200$ workers: $X_1, \dots, X_{200}$.

Possible estimators for μ :

- ① $\hat{\mu}_1 = \bar{X} = \frac{1}{200} \sum_{i=1}^{200} X_i$ (sample mean)
- ② $\hat{\mu}_2 = \tilde{X}$ (sample median)
- ③ $\hat{\mu}_3 = \frac{X_{(1)} + X_{(200)}}{2}$ (midrange)

All three are **estimators**—different recipes. Which one should we use?

Key Question

We need **criteria** to compare estimators. The main criteria: **unbiasedness**, **efficiency**, and **consistency**.

Unbiasedness: Hitting the Target on Average

Definition

An estimator $T = T(X_1, \dots, X_n)$ is **unbiased** for θ if:

$$\mathbb{E}(T) = \theta$$

The **bias** of T is defined as:

$$\text{Bias}(T) = \mathbb{E}(T) - \theta$$

An unbiased estimator does not **systematically** over- or underestimate the true parameter. If you could repeat the sampling experiment infinitely many times, the *average* of all your estimates would equal the true value θ .

Warning: Unbiased does *not* mean every single estimate is close to θ —it means there is no systematic tendency to miss in one direction.

The Height Example: Why Sampling Matters

Goal: Estimate the average height μ of all METU students.

Method 1: Randomly select 50 students from the entire student body and compute $\bar{X}$.

$\Rightarrow \mathbb{E}(\bar{X}) = \mu$ ✓ **Unbiased!**

Method 2: Go to the men's basketball team and measure their heights.

$\Rightarrow \mathbb{E}(\bar{X}_{\text{basketball}}) > \mu$ ✗ **Biased!**

What Went Wrong?

The basketball sample is **not representative**:

- Only males (taller on average)
- Athletes selected for height
- The estimator *systematically* overestimates μ

Bias often comes from **non-random sampling**. In economics, selection bias is everywhere: surveying only employed workers overestimates wages; using only listed firms overestimates firm size.

Proof: $\bar{X}$ Is Unbiased for μ

Theorem

Let $X_1, \dots, X_n$ be i.i.d. with $\mathbb{E}(X_i) = \mu$. Then $\bar{X}$ is unbiased for μ .

Proof:

$$\begin{aligned}\mathbb{E}(\bar{X}) &= \mathbb{E}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}(X_i) && \text{(linearity of expectation)} \\ &= \frac{1}{n} \sum_{i=1}^n \mu = \frac{1}{n} \cdot n\mu = \mu && \checkmark\end{aligned}$$

This proof is short but powerful. The key step is **linearity of expectation**: $\mathbb{E}(\sum) = \sum \mathbb{E}$. This works even without independence!

Why $n - 1$? The Sample Variance

Recall: $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$. **Why divide by $n - 1$ instead of n ?**

Why $n - 1$? The Sample Variance

Recall: $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$. **Why divide by $n - 1$ instead of n ?**

The Problem with Dividing by n

If we define $\tilde{S}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$, then $\mathbb{E}(\tilde{S}^2) = \frac{n-1}{n} \sigma^2 \neq \sigma^2$. The estimator with n **systematically underestimates** σ^2 —it is **biased downward**.

We lose one “degree of freedom” because we used $\bar{X}$ instead of the true μ . Data points are pulled toward $\bar{X}$, making deviations look smaller. Dividing by $n - 1$ corrects this shrinkage.

Proof: S^2 Is Unbiased for σ^2

Useful identity: $\sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n X_i^2 - n\bar{X}^2$

Proof: S^2 Is Unbiased for σ^2

Useful identity: $\sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n X_i^2 - n\bar{X}^2$

Take expectations (using $\mathbb{E}(X_i^2) = \sigma^2 + \mu^2$, $\mathbb{E}(\bar{X}^2) = \sigma^2/n + \mu^2$):

$$\mathbb{E}\left[\sum (X_i - \bar{X})^2\right] = n(\sigma^2 + \mu^2) - n\left(\frac{\sigma^2}{n} + \mu^2\right) = (n-1)\sigma^2$$

Proof: S^2 Is Unbiased for σ^2

Useful identity: $\sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n X_i^2 - n\bar{X}^2$

Take expectations (using $\mathbb{E}(X_i^2) = \sigma^2 + \mu^2$, $\mathbb{E}(\bar{X}^2) = \sigma^2/n + \mu^2$):

$$\mathbb{E}\left[\sum (X_i - \bar{X})^2\right] = n(\sigma^2 + \mu^2) - n\left(\frac{\sigma^2}{n} + \mu^2\right) = (n-1)\sigma^2$$

Therefore:

$$\mathbb{E}(S^2) = \frac{1}{n-1} \cdot (n-1)\sigma^2 = \sigma^2 \quad \checkmark$$

The key insight: $\mathbb{E}(X^2) = \text{Var}(X) + [\mathbb{E}(X)]^2$. We apply this to both X_i and $\bar{X}$, and the μ^2 terms cancel.

Degrees of Freedom: The Intuition

Why exactly $n - 1$?

We have n data points, but used one to compute $\bar{X}$. This leaves $n - 1$ **independent** pieces of information for estimating variance.

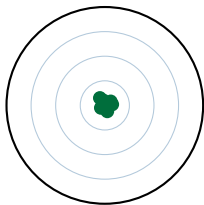
Formally: $X_1 - \bar{X}, \dots, X_n - \bar{X}$ satisfy $\sum (X_i - \bar{X}) = 0$, so only $n - 1$ are free to vary.

Analogy: 5 friends split a 500 TL bill; the first 4 pay freely, but the 5th is fixed: $500 - (\text{sum of 4})$. Similarly, knowing $\bar{X}$ and $n - 1$ deviations fixes the last.

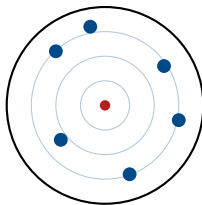
Practical Note

As n grows, $\frac{1}{n-1} \approx \frac{1}{n}$, so the distinction vanishes. It matters most for small samples.

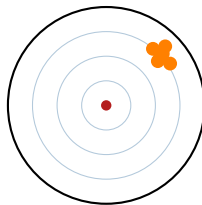
The Dartboard Analogy



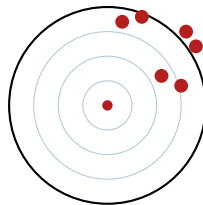
Low Bias, Low Var
IDEAL



Low Bias, High Var
Unbiased but noisy



High Bias, Low Var
Precise but wrong



High Bias, High Var
WORST

Bias = how far the *average* dart lands from the bullseye (systematic error). **Variance** = how *spread out* the darts are (random error). The ideal estimator has **both** low bias **and** low variance.

Efficiency: Among the Unbiased, Choose the Tightest

Definition

Among all **unbiased** estimators of θ , the one with the **smallest variance** is called the **most efficient** (or the **minimum variance unbiased estimator**, MVUE).

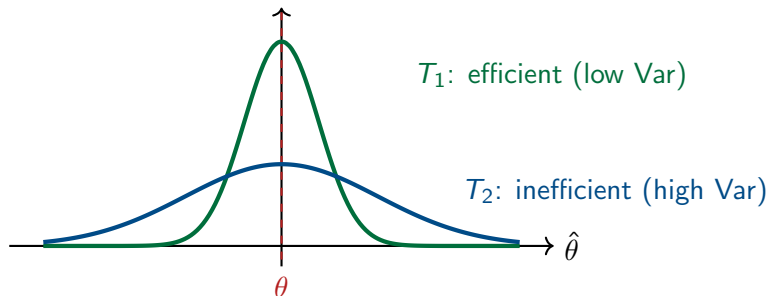
The **relative efficiency** of T_1 relative to T_2 (both unbiased for θ) is:

$$\text{eff}(T_1, T_2) = \frac{\text{Var}(T_2)}{\text{Var}(T_1)}$$

If $\text{eff}(T_1, T_2) > 1$, then T_1 is **more efficient** than T_2 .

If two estimators are both unbiased, they both “aim at the right target.” The more efficient one has a **tighter cluster** around θ . Lower variance means your answer is more likely to be *close* to the truth—in economics, this means more **precise** policy recommendations.

Efficiency: A Visual Comparison



Both T_1 and T_2 are centered on θ (both unbiased), but T_1 is **concentrated** more tightly. With T_1 , any single estimate is more likely to be near θ .

Choosing the more efficient estimator is like choosing a **more precise measuring instrument**. Both give the right answer on average, but one gives you less noise.

Example: Mean vs. Median for Normal Data

Let $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} N(\mu, \sigma^2)$. Both $\bar{X}$ and $\tilde{X}$ (sample median) are unbiased for μ .

Variances:

$$\text{Var}(\bar{X}) = \frac{\sigma^2}{n} \qquad \text{Var}(\tilde{X}) \approx \frac{\pi}{2} \cdot \frac{\sigma^2}{n} \approx \frac{1.571 \sigma^2}{n}$$

Relative efficiency:

$$\text{eff}(\bar{X}, \tilde{X}) = \frac{\text{Var}(\tilde{X})}{\text{Var}(\bar{X})} \approx \frac{\pi}{2} \approx 1.571$$

Conclusion

For normal data, $\bar{X}$ is about **57% more efficient** than the median. You would need roughly 157 observations using the median to get the same precision as 100 using the mean!

However, the median is more **robust** to outliers. With billionaires in income data, the median may be preferable.

Mean Square Error (MSE)

Definition

The **Mean Square Error** of an estimator T for θ is:

$$\text{MSE}(T) = \mathbb{E}[(T - \theta)^2]$$

Theorem (MSE Decomposition)

$$\text{MSE}(T) = \text{Var}(T) + [\text{Bias}(T)]^2$$

MSE measures the **total error** of an estimator: how far T is from θ on average (in squared terms). It combines both random fluctuation (variance) and systematic error (bias). The proof on the next slide shows why this decomposition holds.

Proof of the MSE Decomposition

Goal: Show $\text{MSE}(T) = \text{Var}(T) + [\text{Bias}(T)]^2$. Let $b = \mathbb{E}(T) - \theta$.

Step 1: Insert $\mathbb{E}(T)$: $T - \theta = \underbrace{(T - \mathbb{E}(T))}_{\text{random part}} + \underbrace{(\mathbb{E}(T) - \theta)}_b$

Proof of the MSE Decomposition

Goal: Show $\text{MSE}(T) = \text{Var}(T) + [\text{Bias}(T)]^2$. Let $b = \mathbb{E}(T) - \theta$.

Step 1: Insert $\mathbb{E}(T)$: $T - \theta = \underbrace{(T - \mathbb{E}(T))}_{\text{random part}} + \underbrace{(\mathbb{E}(T) - \theta)}_b$

Step 2: Square and take expectation:

$$\mathbb{E}[(T - \theta)^2] = \mathbb{E}[(T - \mathbb{E}(T))^2] + 2b \cdot \underbrace{\mathbb{E}[T - \mathbb{E}(T)]}_{=0} + b^2$$

Proof of the MSE Decomposition

Goal: Show $\text{MSE}(T) = \text{Var}(T) + [\text{Bias}(T)]^2$. Let $b = \mathbb{E}(T) - \theta$.

Step 1: Insert $\mathbb{E}(T)$: $T - \theta = \underbrace{(T - \mathbb{E}(T))}_{\text{random part}} + \underbrace{(\mathbb{E}(T) - \theta)}_b$

Step 2: Square and take expectation:

$$\mathbb{E}[(T - \theta)^2] = \mathbb{E}[(T - \mathbb{E}(T))^2] + 2b \cdot \underbrace{\mathbb{E}[T - \mathbb{E}(T)]}_{=0} + b^2$$

Step 3: Cross-term vanishes ($\mathbb{E}[T - \mathbb{E}(T)] = 0$), so:

$$\text{MSE}(T) = \underbrace{\mathbb{E}[(T - \mathbb{E}(T))^2]}_{\text{Var}(T)} + b^2 = \text{Var}(T) + [\text{Bias}(T)]^2 \quad \checkmark$$

The key trick: add and subtract $\mathbb{E}(T)$. The cross-term vanishes because $T - \mathbb{E}(T)$ has zero mean. This decomposition is used throughout statistics and econometrics.

The Bias–Variance Tradeoff

$$\text{MSE}(T) = \underbrace{\text{Var}(T)}_{\text{precision}} + \underbrace{[\text{Bias}(T)]^2}_{\text{accuracy}}$$

- **Variance** measures random fluctuation (precision)
- **Bias²** measures systematic error (accuracy)
- These two components often **trade off**: reducing one may increase the other

Why This Matters

Unbiasedness alone is not enough. An estimator with small bias but much lower variance can have smaller MSE—closer to the truth on average. Concrete example on the next slides.

MSE Example: Estimating σ^2

Compare two estimators for σ^2 from a $N(\mu, \sigma^2)$ population:

$$S^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2$$

- Bias = 0 (unbiased)
- $\text{Var}(S^2) = \frac{2\sigma^4}{n-1}$
- $\text{MSE}(S^2) = \frac{2\sigma^4}{n-1}$

$$\tilde{S}^2 = \frac{1}{n} \sum (X_i - \bar{X})^2$$

- Bias = $-\frac{\sigma^2}{n}$ (biased!)
- $\text{Var}(\tilde{S}^2) = \frac{2(n-1)\sigma^4}{n^2}$
- $\text{MSE}(\tilde{S}^2) = \frac{(2n-1)\sigma^4}{n^2}$

Comparison: $\text{MSE}(\tilde{S}^2) = \frac{2n-1}{n^2}\sigma^4 < \frac{2}{n-1}\sigma^4 = \text{MSE}(S^2)$ for $n \geq 2$.

The **biased** estimator $\tilde{S}^2$ has *smaller MSE* than the unbiased S^2 ! Its lower variance more than compensates for the bias. This is a concrete example of the bias–variance tradeoff.

Deriving the Variances: Step by Step

Key fact: $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$, and $Q \sim \chi^2(k) \Rightarrow \text{Var}(Q) = 2k$.

Variance of S^2 :

$$\text{Var}\left(\frac{(n-1)S^2}{\sigma^2}\right) = 2(n-1) \Rightarrow \text{Var}(S^2) = \boxed{\frac{2\sigma^4}{n-1}}$$

Deriving the Variances: Step by Step

Key fact: $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$, and $Q \sim \chi^2(k) \Rightarrow \text{Var}(Q) = 2k$.

Variance of S^2 :

$$\text{Var}\left(\frac{(n-1)S^2}{\sigma^2}\right) = 2(n-1) \Rightarrow \text{Var}(S^2) = \boxed{\frac{2\sigma^4}{n-1}}$$

Variance of $\tilde{S}^2$: Since $\tilde{S}^2 = \frac{n-1}{n}S^2$:

$$\text{Var}(\tilde{S}^2) = \left(\frac{n-1}{n}\right)^2 \text{Var}(S^2) = \boxed{\frac{2(n-1)\sigma^4}{n^2}}$$

MSE of $\tilde{S}^2$: $\frac{2(n-1)\sigma^4}{n^2} + \frac{\sigma^4}{n^2} = \frac{(2n-1)\sigma^4}{n^2} \quad \checkmark$

Consistency: Getting It Right Eventually

Definition

An estimator $T_n = T_n(X_1, \dots, X_n)$ is **consistent** for θ if:

$$T_n \xrightarrow{P} \theta \quad \text{i.e., } \forall \varepsilon > 0 : \quad \lim_{n \rightarrow \infty} \mathbb{P}(|T_n - \theta| \geq \varepsilon) = 0$$

A consistent estimator gets **arbitrarily close** to θ as the sample size grows. With enough data, you can pin down the truth to any desired precision.

Theorem (Sufficient Condition for Consistency)

If $MSE(T_n) \rightarrow 0$ as $n \rightarrow \infty$, then T_n is consistent for θ .

Since $MSE = \text{Var} + \text{Bias}^2$, consistency holds if **both** the variance and the bias vanish as $n \rightarrow \infty$.

Consistency: Two Examples

Example 1: The sample mean $\bar{X}$ for estimating μ .

- $\text{Bias}(\bar{X}) = 0$ (unbiased for all n)
- $\text{Var}(\bar{X}) = \sigma^2/n \rightarrow 0$ as $n \rightarrow \infty$
- $\text{MSE}(\bar{X}) = \sigma^2/n \rightarrow 0 \Rightarrow \bar{X}$ is **consistent** for μ . ✓

Example 2: The biased estimator $\tilde{S}^2 = \frac{1}{n} \sum (X_i - \bar{X})^2$ for estimating σ^2 .

- $\text{Bias}(\tilde{S}^2) = -\sigma^2/n \rightarrow 0$ as $n \rightarrow \infty$
- $\text{Var}(\tilde{S}^2) = \frac{2(n-1)\sigma^4}{n^2} \rightarrow 0$ as $n \rightarrow \infty$
- $\text{MSE}(\tilde{S}^2) \rightarrow 0 \Rightarrow \tilde{S}^2$ is **consistent** for σ^2 . ✓

$\tilde{S}^2$ is biased for every finite n , but the bias shrinks to zero. **Consistency does not require unbiasedness**—it only requires that total error (MSE) vanishes. This is why consistency is considered a *minimum requirement* for any sensible estimator.

Summary of Estimator Properties So Far

Property	What It Means	Formal Criterion
Unbiasedness	Correct on average	$\mathbb{E}(T) = \theta$
Efficiency	Smallest spread (among unbiased)	$\text{Var}(T)$ is minimal
Low MSE	Small total error	$\text{Var}(T) + \text{Bias}^2$ is small
Consistency	Converges to θ	$T_n \xrightarrow{P} \theta$

- **Consistency** = minimum requirement (with enough data, get the right answer)
- **Unbiasedness** = required (no systematic error)
- **Efficiency** = gold standard among unbiased estimators
- **MSE** = most complete single measure (accounts for bias *and* variance)

Example 3.1: Uniform Distribution $U(0, \beta)$

Let $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} U(0, \beta)$ where $\beta > 0$ is unknown.

Population facts: $\mathbb{E}(X_i) = \beta/2$, $\text{Var}(X_i) = \beta^2/12$.

Two candidate estimators for β :

① $\hat{\beta}_1 = 2\bar{X}$ (method of moments: since $\mathbb{E}(X) = \beta/2$, solve for β)

② $\hat{\beta}_2 = \frac{n+1}{n} X_{(n)}$ (adjusted maximum order statistic)

where $X_{(n)} = \max(X_1, \dots, X_n)$.

Both estimators are constructed differently, yet both aim to estimate β . Let us check which one is “best” by computing bias, variance, and MSE for each.

Example 3.1: Computing Bias and Variance

Estimator 1: $\hat{\beta}_1 = 2\bar{X}$

$$\mathbb{E}(\hat{\beta}_1) = 2\mathbb{E}(\bar{X}) = 2 \cdot \frac{\beta}{2} = \beta \quad \Rightarrow \text{Unbiased}$$

$$\text{Var}(\hat{\beta}_1) = 4 \text{Var}(\bar{X}) = 4 \cdot \frac{\beta^2}{12n} = \frac{\beta^2}{3n}$$

Example 3.1: Computing Bias and Variance

Estimator 1: $\hat{\beta}_1 = 2\bar{X}$

$$\mathbb{E}(\hat{\beta}_1) = 2\mathbb{E}(\bar{X}) = 2 \cdot \frac{\beta}{2} = \beta \Rightarrow \text{Unbiased}$$

$$\text{Var}(\hat{\beta}_1) = 4 \text{Var}(\bar{X}) = 4 \cdot \frac{\beta^2}{12n} = \frac{\beta^2}{3n}$$

Estimator 2: $\hat{\beta}_2 = \frac{n+1}{n}X_{(n)}$

Using the pdf of $X_{(n)}$: $f_{X_{(n)}}(x) = \frac{nx^{n-1}}{\beta^n}$ for $0 < x < \beta$:

$$\mathbb{E}(X_{(n)}) = \frac{n}{n+1}\beta, \quad \text{Var}(X_{(n)}) = \frac{n\beta^2}{(n+1)^2(n+2)}$$

$$\mathbb{E}(\hat{\beta}_2) = \frac{n+1}{n} \cdot \frac{n}{n+1}\beta = \beta \Rightarrow \text{Unbiased}$$

$$\text{Var}(\hat{\beta}_2) = \left(\frac{n+1}{n}\right)^2 \text{Var}(X_{(n)}) = \frac{\beta^2}{n(n+2)}$$

Example 3.1: Comparison

Both estimators are unbiased. Their MSEs equal their variances:

	$\hat{\beta}_1 = 2\bar{X}$	$\hat{\beta}_2 = \frac{n+1}{n}X_{(n)}$
Bias	0	0
Variance = MSE	$\frac{\beta^2}{3n}$	$\frac{\beta^2}{n(n+2)}$

Example 3.1: Comparison

Both estimators are unbiased. Their MSEs equal their variances:

	$\hat{\beta}_1 = 2\bar{X}$	$\hat{\beta}_2 = \frac{n+1}{n}X_{(n)}$
Bias	0	0
Variance = MSE	$\frac{\beta^2}{3n}$	$\frac{\beta^2}{n(n+2)}$

For $n = 5$: $\text{MSE}(\hat{\beta}_1) = \beta^2/15 \approx 0.067\beta^2$, $\text{MSE}(\hat{\beta}_2) = \beta^2/35 \approx 0.029\beta^2$.

Relative efficiency: $\text{eff}(\hat{\beta}_2, \hat{\beta}_1) = (n+2)/3$. For $n = 5$: $\text{eff} \approx 2.33$.

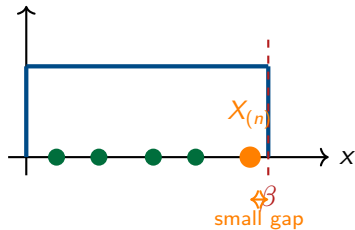
Winner: $\hat{\beta}_2$

$\hat{\beta}_2$ is more efficient—especially for large n , where $\frac{1}{n(n+2)} \ll \frac{1}{3n}$. The maximum carries the most information about the boundary β .

Why Is the Maximum So Informative?

If $X_i \sim U(0, \beta)$, then all observations fall in $[0, \beta]$. The **largest observation** $X_{(n)}$ is always close to β and gets closer as n increases.

The sample mean $\bar{X} \approx \beta/2$, so we must *double* it to estimate β —this amplifies the noise. The maximum directly “touches” the boundary.



Economic analogy: Estimating the maximum willingness to pay in a market. The highest observed bid is more informative than twice the average bid.

Practice Problem 1: Bias and Unbiasedness

Problem: Let $X_1, \dots, X_n$ be i.i.d. with $\mathbb{E}(X_i) = \mu$ and $\text{Var}(X_i) = \sigma^2$. Consider the estimator:

$$T = \frac{1}{n+1} \sum_{i=1}^n X_i$$

- (a) Is T unbiased for μ ? If not, find the bias.
- (b) Compute $\text{MSE}(T)$.
- (c) For what values of μ and σ^2 is $\text{MSE}(T) < \text{MSE}(\bar{X})$?

Practice Problem 1: Bias and Unbiasedness

Problem: Let $X_1, \dots, X_n$ be i.i.d. with $\mathbb{E}(X_i) = \mu$ and $\text{Var}(X_i) = \sigma^2$. Consider the estimator:

$$T = \frac{1}{n+1} \sum_{i=1}^n X_i$$

- (a) Is T unbiased for μ ? If not, find the bias.
- (b) Compute $\text{MSE}(T)$.
- (c) For what values of μ and σ^2 is $\text{MSE}(T) < \text{MSE}(\bar{X})$?

Solution (a):

$$\mathbb{E}(T) = \frac{1}{n+1} \sum_{i=1}^n \mu = \frac{n\mu}{n+1} \neq \mu$$

T is **biased**: $\text{Bias}(T) = \frac{n\mu}{n+1} - \mu = \frac{-\mu}{n+1}$

Practice Problem 1 (Solution Continued)

(b) $\text{Var}(T) = \frac{n\sigma^2}{(n+1)^2}$, so $\text{MSE}(T) = \frac{n\sigma^2 + \mu^2}{(n+1)^2}$.

Practice Problem 1 (Solution Continued)

(b) $\text{Var}(T) = \frac{n\sigma^2}{(n+1)^2}$, so $\text{MSE}(T) = \frac{n\sigma^2 + \mu^2}{(n+1)^2}$.

(c) We need $\text{MSE}(T) < \text{MSE}(\bar{X}) = \sigma^2/n$:

$$\frac{n\sigma^2 + \mu^2}{(n+1)^2} < \frac{\sigma^2}{n} \implies n\mu^2 < (2n+1)\sigma^2 \implies \frac{\mu^2}{\sigma^2} < \frac{2n+1}{n}$$

When $|\mu|/\sigma$ is small (low signal-to-noise), the **biased** estimator T beats the unbiased $\bar{X}$ in MSE! Shrinking toward zero helps when the true mean is close to zero—a manifestation of the bias–variance tradeoff.

Practice Problem 2: Comparing Estimators

Problem: An economist wants to estimate the average GDP growth rate μ using n quarterly observations. Two estimators are proposed:

① $T_1 = \bar{X}$ (sample mean)

② $T_2 = \frac{X_1 + X_n}{2}$ (average of first and last observation)

Assume X_i i.i.d. with mean μ and variance σ^2 .

- (a) Show both are unbiased.
- (b) Which is more efficient?

Practice Problem 2: Comparing Estimators

Problem: An economist wants to estimate the average GDP growth rate μ using n quarterly observations. Two estimators are proposed:

- ① $T_1 = \bar{X}$ (sample mean)
- ② $T_2 = \frac{X_1 + X_n}{2}$ (average of first and last observation)

Assume X_i i.i.d. with mean μ and variance σ^2 .

- (a) Show both are unbiased.
- (b) Which is more efficient?

Solution (a): $\mathbb{E}(T_1) = \mu \checkmark$ and $\mathbb{E}(T_2) = \frac{\mu + \mu}{2} = \mu \checkmark$

Practice Problem 2: Comparing Estimators

Problem: An economist wants to estimate the average GDP growth rate μ using n quarterly observations. Two estimators are proposed:

- ① $T_1 = \bar{X}$ (sample mean)
- ② $T_2 = \frac{X_1 + X_n}{2}$ (average of first and last observation)

Assume X_i i.i.d. with mean μ and variance σ^2 .

- (a) Show both are unbiased.
- (b) Which is more efficient?

Solution (a): $\mathbb{E}(T_1) = \mu \checkmark$ and $\mathbb{E}(T_2) = \frac{\mu + \mu}{2} = \mu \checkmark$

Solution (b):

$$\text{Var}(T_1) = \frac{\sigma^2}{n}, \quad \text{Var}(T_2) = \frac{\sigma^2/1 + \sigma^2/1}{4} = \frac{\sigma^2}{2}$$

Since $\sigma^2/n < \sigma^2/2$ for $n > 2$, T_1 is more efficient.

T_2 **wastes** $n - 2$ observations! **Using all data is almost always better.**

Practice Problem 3: MSE and Consistency

Problem: Let $X_1, \dots, X_n$ be i.i.d. with $\mathbb{E}(X_i) = \mu$ and $\text{Var}(X_i) = \sigma^2$. Consider the estimator $T_n = \frac{n}{n+2} \bar{X}$ for estimating μ .

- (a) Find the bias of T_n .
- (b) Compute $\text{MSE}(T_n)$.
- (c) Is T_n consistent for μ ?

Practice Problem 3: MSE and Consistency

Problem: Let $X_1, \dots, X_n$ be i.i.d. with $\mathbb{E}(X_i) = \mu$ and $\text{Var}(X_i) = \sigma^2$. Consider the estimator $T_n = \frac{n}{n+2} \bar{X}$ for estimating μ .

- (a) Find the bias of T_n .
- (b) Compute $\text{MSE}(T_n)$.
- (c) Is T_n consistent for μ ?

Solution (a): $\mathbb{E}(T_n) = \frac{n}{n+2}\mu$, so $\text{Bias}(T_n) = \frac{n}{n+2}\mu - \mu = \frac{-2\mu}{n+2}$.

Practice Problem 3: MSE and Consistency

Problem: Let $X_1, \dots, X_n$ be i.i.d. with $\mathbb{E}(X_i) = \mu$ and $\text{Var}(X_i) = \sigma^2$. Consider the estimator $T_n = \frac{n}{n+2} \bar{X}$ for estimating μ .

- (a) Find the bias of T_n .
- (b) Compute $\text{MSE}(T_n)$.
- (c) Is T_n consistent for μ ?

Solution (a): $\mathbb{E}(T_n) = \frac{n}{n+2}\mu$, so $\text{Bias}(T_n) = \frac{n}{n+2}\mu - \mu = \frac{-2\mu}{n+2}$.

Solution (b): $\text{Var}(T_n) = \frac{n^2}{(n+2)^2} \cdot \frac{\sigma^2}{n} = \frac{n\sigma^2}{(n+2)^2}$

$$\text{MSE}(T_n) = \frac{n\sigma^2}{(n+2)^2} + \frac{4\mu^2}{(n+2)^2} = \frac{n\sigma^2 + 4\mu^2}{(n+2)^2}$$

Practice Problem 3: MSE and Consistency

Problem: Let $X_1, \dots, X_n$ be i.i.d. with $\mathbb{E}(X_i) = \mu$ and $\text{Var}(X_i) = \sigma^2$. Consider the estimator $T_n = \frac{n}{n+2} \bar{X}$ for estimating μ .

- (a) Find the bias of T_n .
- (b) Compute $\text{MSE}(T_n)$.
- (c) Is T_n consistent for μ ?

Solution (a): $\mathbb{E}(T_n) = \frac{n}{n+2}\mu$, so $\text{Bias}(T_n) = \frac{n}{n+2}\mu - \mu = \frac{-2\mu}{n+2}$.

Solution (b): $\text{Var}(T_n) = \frac{n^2}{(n+2)^2} \cdot \frac{\sigma^2}{n} = \frac{n\sigma^2}{(n+2)^2}$

$$\text{MSE}(T_n) = \frac{n\sigma^2}{(n+2)^2} + \frac{4\mu^2}{(n+2)^2} = \frac{n\sigma^2 + 4\mu^2}{(n+2)^2}$$

Solution (c): As $n \rightarrow \infty$: $\text{Bias} = \frac{-2\mu}{n+2} \rightarrow 0$ and $\text{Var} = \frac{n\sigma^2}{(n+2)^2} \rightarrow 0$.

So $\text{MSE}(T_n) \rightarrow 0$, meaning T_n is **consistent** for μ . ✓

Practice Problem 4: Comparing Estimators

Problem: For estimating θ , you have three estimators:

	T_A	T_B	T_C
$\mathbb{E}(T)$	θ	$\theta + 2$	θ
$\text{Var}(T)$	25	10	36

Compare these estimators for **efficiency** and **MSE**.

Practice Problem 4: Comparing Estimators

Problem: For estimating θ , you have three estimators:

	T_A	T_B	T_C
$\mathbb{E}(T)$	θ	$\theta + 2$	θ
$\text{Var}(T)$	25	10	36

Compare these estimators for **efficiency** and **MSE**.

Solution: $\text{MSE}(T_A) = 25$, $\text{MSE}(T_B) = 10 + 4 = 14$ (**smallest!**), $\text{MSE}(T_C) = 36$.

Efficiency: Among unbiased (T_A , T_C), T_A is more efficient ($\text{Var} = 25 < 36$). **MSE:** T_B has smallest MSE despite being biased—low variance compensates.

Key Takeaways from Week 3

- 1 An **estimator** is a random variable (a rule); an **estimate** is a number (a realization)
- 2 **Unbiasedness** ($\mathbb{E}(T) = \theta$): no systematic error—but a single estimate can still be far off
- 3 $\bar{X}$ is unbiased for μ ; S^2 (with $n - 1$) is unbiased for σ^2
- 4 We divide by $n - 1$ because we lose one **degree of freedom** when estimating μ by $\bar{X}$
- 5 **Efficiency**: among unbiased estimators, prefer the one with *smallest variance*
- 6 $\text{MSE} = \text{Variance} + \text{Bias}^2$: a biased estimator can beat an unbiased one!
- 7 **Consistency**: $T_n \xrightarrow{P} \theta$ as $n \rightarrow \infty$ —the minimum requirement for a sensible estimator

The Hierarchy

Best estimator = Low MSE = Good balance of bias and variance.

Unbiasedness alone is *not enough*—always consider variance and MSE too.

A Recap: The Big Picture

So far, we have learned about some desired properties of point estimators. We will keep in mind that we want to have unbiased estimators with minimum variance so that we have more precise estimates or predictions of the unknown population parameters. But when we can't have an unbiased estimator, we at least want an estimator that is consistent (i.e. converges in probability to the population parameter). This is why I recommend you to try your best to understand the basic theorems such as the Weak Law of Large Numbers, Chebyshev's Theorem and Central Limit Theorem which help us find unbiased or consistent estimators. You should also make sure that you are comfortable with taking expectations, and probability limits.

Further Reading

- Wackerly, Mendenhall & Scheaffer, Ch. 8: Estimation
- Hogg, McKean & Craig, Ch. 6: Point Estimation
- Casella & Berger, Ch. 7: Point Estimation (advanced)
- Devore & Berk, Ch. 6: Point Estimation

Key terms to review: estimator, estimate, bias, unbiasedness, efficiency, relative efficiency, mean square error, bias–variance tradeoff, consistency, degrees of freedom.